Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 2 February 2008 (MN LaTeX style file v2.2)

Conditional Cumulants in Weakly Non-linear Regime

Jun Pan⋆ and István Szapudi†
Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu HI 96822, USA

2 February 2008

ABSTRACT

Conditional cumulants form a set of unique statistics which represent a sensible compromise between N-point correlation functions and cumulants measured from moments of counts in cells. They share accurate edge corrected estimators with N-point correlation functions, yet, they are as straightforward to measure and interpret as counts in cells. The conditional cumulants have three equivalent views as i) degenerate N-point correlation functions ii) or integrated monopole moments of the bispectrum iii) they are closely related to neighbour counts. We compute the predictions of weakly non-linear perturbation theory for conditional cumulants and compare them with measurements in simulations, both in real and redshift space. We find excellent agreement between theory and simulations, especially on scales $\gtrsim 20\mathrm{h}^{-1}\mathrm{Mpc}$. Due to their advantageous statistical properties and well understood dynamics, we propose conditional cumulants as tools for high precision cosmology. Potential applications include constraining bias and redshift distortions from galaxy redshift surveys.

Key words: Cosmology: theory – large scale structure of the Universe – Methods : statistical

1 INTRODUCTION

Large scale structure statistics of higher than second order contain a wealth of information on cosmological parameters, gravitational amplification of initial fluctuations, and structure formation in general. In particular, higher order statistics have the potential to provide some of the best constraints on the phenomenon of "biasing" (Kaiser 1984). While the core ideas have been known for over two decades (cf. Peebles 1980), the latest wide field galaxy surveys, such as the Sloan Digital Sky Survey (e.g., York et al. 2000, SDSS), and the 2 degree Filed Survey (Colless et al. 2001, 2dF) have motivated a concerted effort to enhance theories and techniques of higher order statistics to the level of "high precision cosmology" (for summary, see Bernardeau et al. 2002).

Counts in cells (CIC) and related statistics have the most well understood theoretical background and consequently yielded some of the most successful applications (e.g., Gaztañaga & Frieman 1994; Szapudi, Meiksin & Nichol 1996; Szapudi et al. 2002). The estimation methods and the corresponding errors have been worked out in detail (e.g. Szapudi & Colombi 1996; Szapudi et al. 1999). But when CIC are measured to constrain theories at a few percent level, edge effects present a major problem due to the complex geometry and cut out holes of realistic surveys. As shown by Szapudi & Colombi (1996), edge effects cannot be corrected for exactly; an approximate estimator exists if the shape dependence of counts is weak (Szapudi 1998b).

On the other hand, full edge effect correction is possible when using the class of estimators by Szapudi & Szalay (1998) for the N-point correlation functions. N-point correlation functions are, however, inherently more complicated objects than cumulants of CIC. They depend on a large number of parameters, their measurement is computationally intensive (Moore et al. 2001), and consequently, their interpretation is difficult. Much of the theoretical effort has been concentrated on the three-point function, yet weakly non-linear perturbation theory and halo models have only limited success when contrasted with simulations (Barriga & Gaztañaga 2002; Takada & Jain 2003). Redshift distortions present an even more formidable challenge.

The goal of the present paper is to explore a set of statistics, the conditional cumulants, which combine the simplicity and transparency of cumulants with the edge effect corrected estimator of the N-point correlation. They can be viewed as degenerate N-point correlation functions, or integrals of the monopole moment of the bispectrum (Szapudi 2004). They are also closely related to (factorial) moments of neighbour counts (Peebles 1980). While previously known, neighbour counts have been used fairly infrequently in comparison to cumulants (e.g., Borgani 1995). The terminology

⋆ jpan@ifa.hawaii.edu
† szapudi@ifa.hawaii.edu

conditional cumulants was introduced by Bonometto et al. (1995). They have used an estimator based on moments of neighbour counts and developed a theory under the assumptions of stable clustering and scale invariance.

In §2, we present formal definition of the conditional cumulants and their relevant properties. Predictions of the third order conditional cumulant in the weakly non-linear regime are described in §3. In §4, we adapt the edge corrected estimator by Szapudi & Szalay (1998), and compare predictions with measurements in simulations. In §5, we summarise results, present the theory in redshift space together with simulation results, and discuss implications for estimating bias.

2 CONDITIONAL CUMULANTS

Conditional cumulants are defined as the joint connected moment of one unsmoothed and $N-1$ smoothed density fluctuation fields. They are realised by integrals of the N-point correlation function through $N-1$ spherical top-hat windows,

$$U_N(r_1,\dots,r_{N-1}) = \int \xi_N(s_1,\dots,s_{N-1},0) \prod_{i=1}^{N-1} d^3s_i \frac{W_{r_i}(s_i)}{V_i} \quad (1)$$

where $V_i = 4\pi/3r_i^3$ is the volume of the window function W_{r_i}. In the most general case, each top hat window might have a different radius. Further simplification arises if all the top hats are the same, i.e. we define $U_N(r)$ with $r_1 = \dots = r_{N-1} = r$ as the *conditional cumulant* (cf. Bonometto et al. 1995). The U_N subtly differs from the usual cumulant of smoothed field $\overline{\xi}_N$ by one less integral over the window function.

The second order, U_2, is equivalent to the confusingly named J_3 integral (e.g., Peebles 1980),

$$U_2(r) = \frac{3}{r^3} J_3(r) = \frac{1}{(2\pi)^3} \int P(k) w(kr) 4\pi k^2 dk \ , \quad (2)$$

where $w(kr) = 3(\sin kr - kr\cos kr)/(kr)^3$ is the Fourier transform of $W_r(s)$, and $P(k)$ is the power spectrum.

For higher orders, we can construct reduced conditional cumulants as

$$R_N(r) = \frac{U_N(r)}{U_2^{N-1}(r)} \ . \quad (3)$$

U_N and R_N have deep connection with moments of neighbour counts (e.g. Peebles 1980). Let us define the partition function $Z[J] = \langle \exp(\int iJ\rho)\rangle$ (cf. Szapudi & Szalay 1993), where ρ is the smoothed density field. Then we can use the special source function $iJ(x) = W(x)s + \delta_D(x)t$ to obtain the generating function $G(s,t)$. This is related to the generating function of neighbour counts factorial moments as $G(s) = \partial_t G(s,t)|_{t=0}$. The final result is

$$G(s) = \sum_{M\geq 1} \frac{(snV)^M}{M!} U_{M+1} \exp \sum_{N\geq 1} \frac{(snV)^N}{N!} \overline{\xi}_N \ , \quad (4)$$

where $nV = \bar{N}$ is the average count of galaxies, and $\overline{\xi}_1 = U_1 = 1$ by definition. This generating function can be used to obtain U_N's and/or R_N's from neighbour counts factorial moments analogously as the generating functions in Szapudi & Szalay (1993) for obtaining S_N's from factorial moments of CIC. For completeness, the generating function for neighbour counts distribution is obtained by substituting $s \to s-1$, while the ordinary moment generating function by $s \to e^s - 1$. We checked that the we recover formulae of Peebles (1980), §36 from $G(e^s-1)$. The above generating function facilitates the extraction of U_N from neighbour counts statistics. For details see Szapudi & Szalay (1993): the entire theory for CIC can be adapted to neighbour counts. So far our discussion has been general; in what follows we will focus on the first non-trivial conditional cumulant U_3.

$U_3(r_1,r_2)$ is simply related to bispectrum by

$$U_3(r_1,r_2) = \frac{1}{(2\pi)^6} \int B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) \\ w(k_1r_1)w(k_2r_2)d^3k_1d^3k_2d^3k_3 \ , \quad (5)$$

where δ_D is the Dirac delta function. To further elucidate the above relation, we use the multi-pole expansion of the bispectrum and the three point correlation function proposed by (Szapudi 2004)

$$B(k_1,k_2,\theta) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} B_l(k_1,k_2) P_l(\cos\theta) \ ; \\ \zeta(r_1,r_2,\theta) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} \zeta_l(r_1,r_2) P_l(\cos\theta) \ , \quad (6)$$

where $\cos\theta$ is $\mathbf{k}_1\cdot\mathbf{k}_2/(k_1k_2)$ or $\mathbf{r}_1\cdot\mathbf{r}_2/(r_1r_2)$, and P_l are Legendre polynomials. The multi-pole moments can be obtained as $B_l = 2\pi\int BP_l d\cos\theta$, $\zeta_l = 2\pi\int \zeta P_l d\cos\theta$. Substituting into the general equation, we find

$$U_3(r_1,r_2) = \frac{4\pi}{V_1V_2} \int_0^{r_1}\int_0^{r_2} \zeta_0(r_1,r_2) r_1^2 r_2^2 dr_1 dr_2 \\ = \frac{4\pi}{(2\pi)^6} \int dk_1 dk_2 \frac{3k_1}{r_1} \frac{3k_2}{r_2} j_1(k_1r_1) j_1(k_2r_2) B_0(k_1,k_2) \ , \quad (7)$$

in which j_1 is the first order spherical Bessel function. We see the U_3 depends only on the monopole moment of the bispectrum/three-point correlation function. This property significantly simplifies the transformation of the statistics under redshift distortions.

3 U_3 IN WEAKLY NON-LINEAR REGIME

On large scales, where the fluctuations are reasonably small, clustering of cosmic structures can be understood in Eulerian weakly non-linear perturbation theory (EPT) (Bernardeau et al. 2002, and references therein). To predict the behaviour of U_3 from Gaussian initial conditions, we assume an expansion of the density field into first, second, etc. order, $\delta = \delta^{(1)} + \delta^{(2)} + \dots$. Then EPT can be used to calculate the leading order contribution to $U_3 = \langle \delta(0)\delta_r^2\rangle_c$, where δ_r is the density field filtered at the scale r, and $\langle\rangle_c$ means connected moment. We use the second-order EPT kernel (Goroff et al. 1986)

$$F_2(\mathbf{k},\mathbf{k}') = \frac{10}{7} + \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\left(\frac{k}{k'}+\frac{k'}{k}\right) + \frac{4}{7}\left(\frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)^2 \ , \quad (8)$$

the linear power spectrum $P(k)$, and integrals of the kernel multiplied with the top-hat window function (Bernardeau

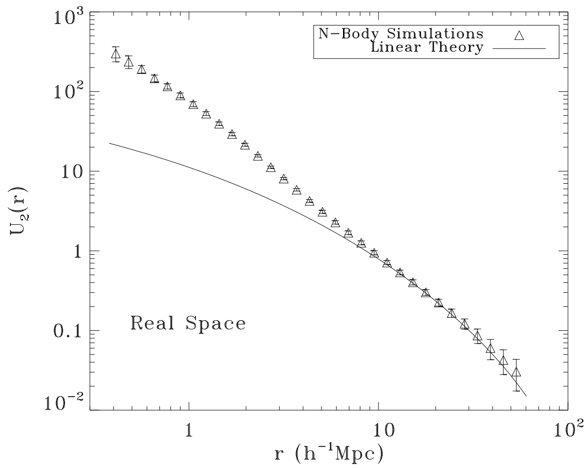

Figure 1. Predictions of $U_2(r)$ on large scales in real space (solid line) compared with measurements in N-body simulations (triangles with error-bars) with ΛCDM cosmology.

1994a) to finally obtain

$$\begin{aligned} R_3(r_1,r_2) \equiv& \frac{U_3(r_1,r_2)}{U_2(r_1)U_2(r_2)} = \frac{34}{21}\left[1+\frac{\overline{\xi}(r_1,r_2)}{U_2(r_1)}+\frac{\overline{\xi}(r_1,r_2)}{U_2(r_2)}\right] \\ &+\frac{1}{3}\frac{\overline{\xi}(r_1,r_2)}{U_2(r_1)}\left[\frac{d\ln U_2(r_2)}{d\ln r_2}+\frac{\partial \ln \overline{\xi}(r_1,r_2)}{\partial \ln r_2}\right] \\ &+\frac{1}{3}\frac{\overline{\xi}(r_1,r_2)}{U_2(r_2)}\left[\frac{d\ln U_2(r_1)}{d\ln r_1}+\frac{\partial \ln \overline{\xi}(r_1,r_2)}{\partial \ln r_1}\right] , \end{aligned} \tag{9}$$

in which $\overline{\xi}(r_1,r_2) = \frac{1}{2\pi^2}\int k^2 P(k) w(kr_1) w(kr_2) dk$. The special case when $r_1 = r_2 = r$ reads

$$R_3 = \frac{34}{21}\left[1+2\frac{\sigma^2}{U_2}\right]+\frac{1}{3}\,\frac{\sigma^2}{U_2}\left[2\frac{d\ln U_2}{d\ln r}+\frac{d\ln\sigma^2}{d\ln r}\right] , \tag{10}$$

where $\sigma^2 = \frac{1}{2\pi^2}\int k^2 P(k) w^2(kr) dk$. The above equations constitute the main results of this paper. Note the similarity of R_3 with the skewness, which is calculated in weakly non-linear perturbation theory as $S_3 = 34/7 - d\ln\sigma^2/d\ln r$ (Juszkiewicz, Bouchet & Colombi 1993; Bernardeau 1994b).

4 MEASUREMENTS

$U_n(r)$ can be measured similarly to N-point correlation functions. For instance U_2 can be thought of as a two point correlation function in a bin $[r_{lo}, r_{hi}] \equiv [0, r]$. Taking the lower limit to be a very small number instead of 0, one can correct for discreteness effects due to self counting (this is equivalent to using factorial moments when neighbour counts are calculated directly). Given a set of data and random points, the class of estimators of Szapudi & Szalay (1998) will provide an edge (and incompleteness) corrected technique to measure conditional cumulants

$$\widehat{U}_n = \frac{(D-R)^n}{R^n} . \tag{11}$$

Existing N-point correlation function codes can be used for the estimation; for higher then third order, one also has to take connected moments in the usual way.

While the above suggests a scaling similar to N-point correlation functions, the relation to neighbour count factorial moments outlined in the previous section can be used to realise the estimator using two-point correlation function codes. To develop such an estimator, neighbour count factorial moments need to be collected for each possible combinations where data and random points play the role of centre and neighbour.

Note that the edge correction of Eq. (11) is expected to be less accurate for conditional cumulants than for N-point correlation functions, however, the estimator will be more accurate than CIC estimators. Several alternative ways for correcting edge effects are known, which would be directly applicable to conditional cumulants (Ripley 1988; Kerscher 1999; Pan & Coles 2002). In what follows, we use Eq. (11) for all results presented. Future high precision measurements could benefit from a shootout of possible estimators as in Kerscher (1999).

To test our theory, we performed measurements in ΛCDM simulations by the Virgo Supercomputing Consortium. We used outputs of the Virgo simulation and the VLS (Very Large Simulation). Except for box sizes and number of particles, these two simulations have identical cosmology parameters: $\Omega_m = 0.3$, $\Omega_v = 0.7$, $\Gamma = 0.21$, $h = 0.7$ and $\sigma_8 = 0.9$. In order to estimate measurement errors, we divide the VLS simulation into eight independent subsets each with the same size and geometry the original Virgo simulation. In total, we have used the resulting nine realisations to estimate errors. Note that we corrected for cosmic bias by always taking the average before ratio statistics were formed.

Our measurements of the second and third order conditional cumulants are displayed in Figs. 1 and 2, respectively. Results from EPT (Eq. 10) are denoted with solid lines. The measurements in simulations are in excellent agreement with EPT, especially on on large scales $\gtrsim 20\mathrm{h}^{-1}$Mpc.

5 SUMMARY AND DISCUSSION

We presented the theory of conditional cumulants in the weakly non-linear regime. The unique set of statistics can be thought of as degenerate N-point correlation functions, or integrated monopole. We have derived the generating function of neighbour count factorial moments, revealing a deep connection to conditional cumulants. We introduced the reduced quantity R_N, which is analogous to the cumulant S_N. We calculated leading order perturbation theory predictions, and showed that results are similar to that of the S_N's. However, while edge correction for CIC is not feasible, we proposed an accurate edge-corrected estimation method for the conditional cumulants. This was applied to a set of measurements in simulations, which yielded results in excellent agreement with the theory, especially on large scales $\gtrsim 20\mathrm{h}^{-1}$Mpc. The agreement with theory encourages further development of this statistic for high precision cosmological applications, such as constraining bias.

As three-dimensional galaxy catalogues are produced inherently in redshift space, understanding effects of redshift distortions on these statistics is crucial before practical applications can follow. In the distant observer approximation, the formula by Kaiser (1987) and Lilje & Efstathiou (1989) is expected to provide an excellent approximation for $U_2(r)$. According to §2, we only need to consider the

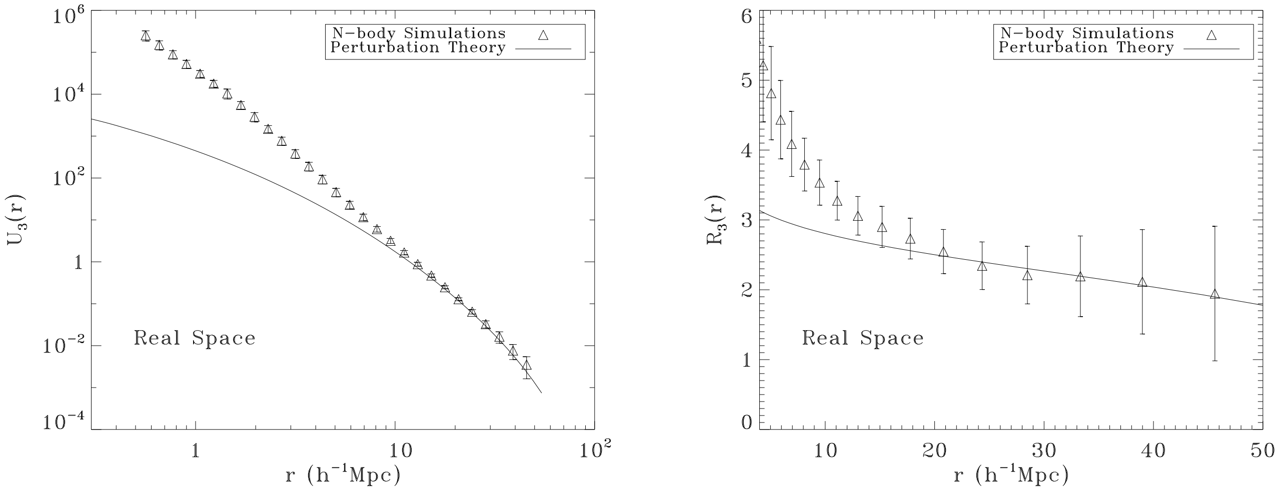

Figure 2. Same as Figure 1. for the third order conditional cumulant U_3 (left), and the reduced statistics R_3 (right).

monopole enhancement

$$U_2(s) = \left(1 + \frac{2}{3}f + \frac{1}{5}f^2\right) U_2(r) \ , \qquad (12)$$

where $f \approx \Omega_m^{0.6}$. This formula essentially predicts a uniform shift of the real space results. To test it, we repeated our measurements in redshift space, and found that the above is indeed an excellent approximation in redshift space (Fig. 3).

Considering the relatively simple, monopole nature of the statistics, we expect that the overall effect on U_3 should also be a simple shift, similarly to the Lagrangian calculations by Hivon et al. (1995) and the more general Eulerian results by Scoccimarro, Couchman & Frieman (1999). Specifically, we propose that ratio of R_3 in redshift space to that in real space can be approximated by

$$\frac{5(2520 + 3360f + 1260f^2 + 9f^3 - 14f^4)}{98(15 + 10f + 3f^2)^2} \cdot \frac{7}{4} \quad . \qquad (13)$$

This is motivated by the notion that the shift from redshift distortions of equilateral triangles should be similar to the corresponding shift for our monopole statistic. Our simulations results (see Fig. 3) show that this simple idea is indeed a surprisingly good approximation, although the phenomenological theory based on the above formula appears to have $\simeq 5\%$ bias on scales $\gtrsim 20\mathrm{h}^{-1}$Mpc where we expect that weakly non-linear perturbation theory is a good approximation. For practical applications, this bias can be calibrated by N-body, or 2LPT (Scoccimarro 2000) simulations.

In addition to the above simple formula, we have calculated the shift due to redshift distortions by angular averaging the bispectrum monopole term in Scoccimarro, Couchman & Frieman (1999). We have found that the results over-predict redshift distortions, however, they would agree with simulations at the 1-2% level if we halved the terms classified as FOG (finger of god). At the moment there is no justification for such a fudge factor, therefore we opt to use the above phenomenology, which is about 5% accurate. While redshift distortions of third order statistics are still not fully understood due to the non-perturbative nature of the redshift space mapping (R. Scoccimarro, private communication), detailed calculations taking into account velocity dispersion effects will improve the accuracy of the redshift space theory U_3.

For applications to constrain bias, one has to keep in mind that redshift distortions and non-linear bias do not commute. However, at the level of the above simple theory, it is clear that one can understand the important effects at least for the third order statistic. There are several ways to apply conditional cumulants for bias determination, either with combination with one other statistic (CIC or cumulant correlators, cf. Szapudi 1998a), or using the configuration dependence of the more general $R_3(r_1, r_2)$. One also has to be careful that in practical applications ratio statistics will contain cosmic bias (Szapudi et al. 1999). We propose that joint estimation with U_2 and U_3 will be more fruitful, even if R_3 is better for plotting purposes. Details of the techniques to constrain bias from these statistics, as well we determination of the bias from wide field redshift surveys is left for future work.

Another way to get around redshift distortions is to adapt conditional cumulants for projected and angular quantities. Such calculations are straightforward, and entirely analogous to those performed for S_3 in the past. Another possible generalisation of our theory would be to use halo models (Cooray & Sheth 2002) to extend the range of applicability of the theory well below $20\mathrm{h}^{-1}$Mpc. These generalisations are left for subsequent research.

ACKNOWLEDGEMENT

This work was supported by NASA through grants AISR NAG5-11996, ATP NAG5-12101, and by NSF through grants AST02-06243 and ITR 1120201-128440. We thank Pablo Fosalba for stimulating discussions. The simulations in this paper were carried out by the Virgo Supercomputing Consortium using computers based at Computing Centre of

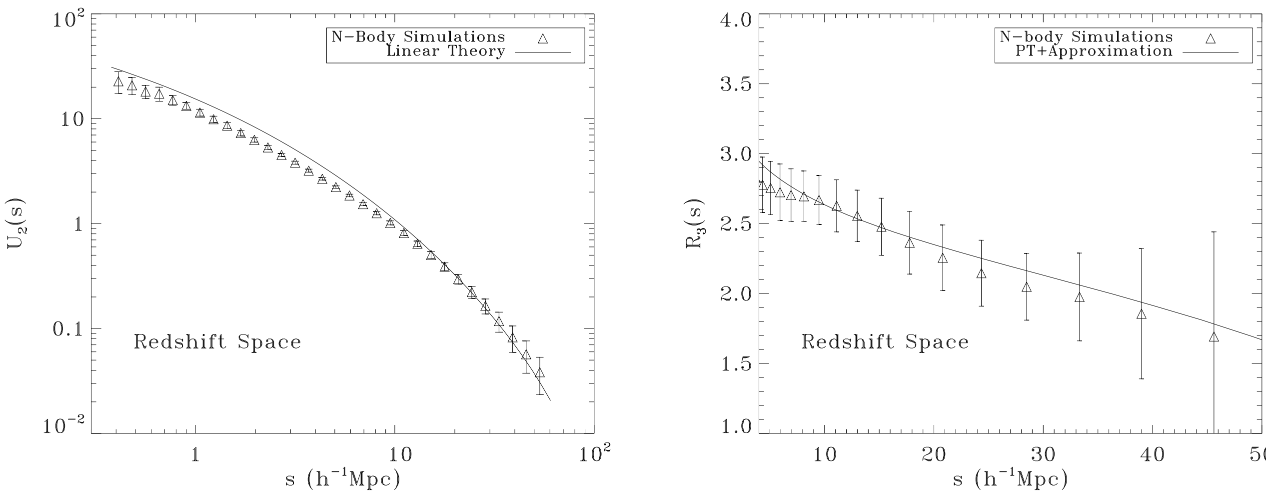

Figure 3. U_2 and R_3. The solid line in left panel comes from Eq. 12. In the right panel the solid line shows phenomenological model based on Eq. 13, the theory appears to be a reasonable approximation at the 5% level.

the Max-Planck Society in Garching and at the Edinburgh Parallel Computing Centre[1].

REFERENCES

Barriga J., Gaztañaga E., 2002, MNRAS, 333, 443
Bernardeau F., 1994a, ApJ, 433, 1
Bernardeau F., 1994b, A&A, 291, 697
Bernardeau F., Colombi S., Gaztanaga E., & Scoccimarro, R., 2002, Phys. Rep., 367, 1
Bonometto S. A., et al., 1995, MNRAS, 273, 101
Borgani S., 1995, Phys. Rep., 251, 1
Colless M., et al., 2001, MNRAS, 328, 1039
Cooray A., & Sheth R., 2002, Phys. Rep., 372, 1
Gaztañaga E., & Frieman J. A., 1994, ApJ, 437, 13
Goroff M. H., et al., 1986, ApJ, 311, 6
Hivon E., Bouchet F. R., Colombi S., & Juszkiewicz, R., 1995, A&A, 298, 643
Juszkiewicz R., Bouchet F R, Colombi S., 1993, ApJ, 412, 9
Kaiser N., 1984, ApJ, 284, 9
Kaiser N., 1987, MNRAS, 227, 1
Kerscher M., 1999, A&A, 343, 333
Lilje Per B., Efstathiou G., 1989, MNRAS, 236, 851
Moore A., et al., 2001, p.71 in *Mining the Sky, Proceedings of the MPA/ESO/MPE WorkShop held at Garching, Germany in 2000*, Springer-Verlag, Heidelberg (astro-ph/0012333)
Pan J., Coles P., 2002, MNRAS, 330, 719
Ripley B. D., 1988, *Statistical inference for spatial processes*, University Press, Cambridge
Peebles P. J. E., 1980, *The Large Scale Struture in Universe*, Princeton Press, Princeton
Scoccimarro R., 2000, ApJ, 544, 597
Scoccimarro R., Couchman H. M. P., & Frieman, H.A., 1999, ApJ, 517, 531
Szapudi I., 1998a, MNRAS, 300, 35
Szapudi I., 1998b, ApJ, 497, 16
Szapudi I., 2004, ApJ, 605, 89
Szapudi I., Colombi S., 1996, ApJ, 470, 131
Szapudi I., Colombi S., Bernardeau F., 1999, MNRAS, 310, 428
Szapudi I., Meiksin A., Nichol R. C., 1996, ApJ, 473, 15
Szapudi I., Szalay A. S., 1993, ApJ, 408, 43
Szapudi I., Szalay A. S., 1998, ApJ, 494, 41
Szapudi I., et al., 2002, ApJ, 570, 75
Takada M., Jain B., 2003, MNRAS, 340, 580
York D.G., et al., 2000, AJ, 120, 1579

[1] The data are publicly available at http://www.mpa-garching.mpg.de/Virgo